Journal of Intelligent Strategic Management
Vol(4), 47-68.
Bumara.ir

Fraud Detection in Supplementary Health Insurance Using Neural Network Algorithms

Masoumeh Esmaeili — PhD in Information Technology Management, Smart Business Branch, Department of Management, Faculty of Management, Islamic Azad University, Kish Branch, Kish, Iran.

Mohammad Malekinia * — Faculty Member, Department of Management, Faculty of Management, Islamic Azad University, South Tehran, Tehran, Iran.

Alireza Pour Ebrahimi — Assistant Professor, Department of Industrial Management, Faculty of Management, Islamic Azad University, Karaj Branch, Karaj, Iran.

Original Research
Received: 2025/01/26
Review: 2025/07/22
Accepted: 2025/09/23
eISSN: 2891-4128

Abstract

Fraud in supplementary health insurance claims is a major challenge that leads to significant financial losses and undermines public trust. In this study, a comprehensive framework for fraud detection using neural network algorithms is presented. This framework utilizes three neural network architectures: a) Multilayer Perceptron (MLP), b) Deep Neural Network (DNN) implemented with Keras, and c) Long-Term Memory Network (LSTM). Raw data were collected from various sources including insurance policy information, insureds, claim files, disease lists, and branch information and then preprocessed with cleaning, transformation, normalization, and outlier removal (using the interquartile range or IQR method). The models were evaluated using the K-means cross-validation method and metrics such as accuracy, confusion matrix, precision, recall, F1 score, and ROC-AUC. Experimental results showed that all models have very high accuracy (about 99.9%), proving that neural network-based systems are capable of reliably distinguishing between legitimate and fraudulent claims. The proposed system can pave the way for more effective real-time fraud monitoring in the insurance industry. These models had very high accuracy in detecting fraud and can be effectively applied in practical and real-world applications. The results of this research indicate the high ability of these models to process complex and nonlinear data, and these algorithms can be used as effective solutions for fraud detection in complementary health insurance and other similar systems. Consequently, this research is an important step towards the development of intelligent models for fraud detection in insurance systems.

Keywords: Fraud detection, complementary therapy, insurance, neural network, data mining, perceptron neural network, deep belief neural network, recurrent neural network

How to Cite: Esmaeili, M. , Malekinia, M. and Pour Ebrahimi, A. (2025). Fraud Detection in Supplementary Health Insurance Using Neural Network Algorithms. Journal of Intelligent Strategic Management .4(3), 47-68.

doi: bumara .3.2.15564.35887873.63081087

Corresponding Author : M_malekinia@azad.ac.ir *

نشریه مدیریت استراتژیک هوشمند
دوره ۴، شماره ۳، پاییز، ۱۴۰۴، ۶۸- ۴۷
Bumara.ir

تشخیص تقلب در بیمه درمان تکمیلی با استفاده از الگوریتم‌های شبکه عصبی

معصومه اسماعیلی دانش آموخته دکتری مدیریت فناوری اطلاعات شاخه کسب و کار هوشمند، گروه مدیریت، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد کیش، کیش، ایران.

محمد ملکی نیا* عضو هیات علمی گروه مدیریت، دانشکده مدیریت، دانشگاه آزاد اسلامی تهران جنوب، تهران، ایران.

علیرضا پورابراهیمی استادیار گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه آزاد اسلامی واحد کرج، کرج، ایران.

مقاله پژوهشی
تاریخ دریافت: ۱۴۰۳/۱۱/۰۷
بازنگری: ۱۴۰۴/۰۴/۳۱
تاریخ پذیرش: ۱۴۰۴/۰۷/۰۱
eISSN: 2891-4128

چکیده

تقلب در مطالبات بیمه درمان تکمیلی، چالشی بزرگ است که منجر به زیان‌های مالی قابل توجه و تضعیف اعتماد عمومی می‌گردد. در این پژوهش، یک چارچوب جامع برای شناسایی تقلب با استفاده از الگوریتم‌های شبکه عصبی ارائه شده است. این چارچوب از سه معماری شبکه عصبی بهره می‌برد: الف) پرسپترون چندلایه(MLP) ، ب) شبکه عصبی عمیق (DNN) پیاده‌سازی‌شده باKeras ، و ج) شبکه حافظه بلندمدت .(LSTM) داده‌های خام از منابع مختلف شامل اطلاعات بیمه‌نامه‌ها، بیمه‌شدگان، پرونده‌های خسارت، فهرست بیماری‌ها و اطلاعات شعب جمع‌آوری و پس از آن با مراحل پاک‌سازی، تبدیل، نرمال‌سازی و حذف داده‌های پرت (با استفاده از روش فاصله بین چارکی یا IQR) پیش‌پردازش شدند. مدل‌ها با استفاده از روش اعتبارسنجی متقاطع-K تایی و معیارهایی نظیر دقت، ماتریس سردرگمی، دقت(Precision) ، بازخوانی(Recall) ، امتیاز F۱ و ROC-AUC ارزیابی شدند. نتایج تجربی نشان دادند که همه مدل‌ها دقت بسیار بالایی (در حدود ۹۹,۹٪) دارند و این موضوع اثبات می‌کند که سیستم‌های مبتنی بر شبکه عصبی قادر به تفکیک قابل‌اعتماد میان ادعاهای مشروع و جعلی هستند. این سیستم پیشنهادی می‌تواند مسیر را برای نظارت لحظه‌ای مؤثرتر بر تقلب در صنعت بیمه هموار کند. این مدل‌ها دقت بسیار بالایی در شناسایی تقلب‌ها داشتند و می‌توانند در کاربردهای عملی و دنیای واقعی به‌طور مؤثر به کار گرفته شوند. نتایج این تحقیق نشان‌دهنده توانایی بالای این مدل‌ها در پردازش داده‌های پیچیده و غیرخطی است و این الگوریتم‌ها می‌توانند به عنوان راه‌حل‌های مؤثری برای تشخیص تقلب در بیمه‌های درمان تکمیلی و سایر سیستم‌های مشابه به کار روند. در نتیجه، این تحقیق گامی مهم در جهت توسعه مدل‌های هوشمند برای شناسایی تقلب در سیستم‌های بیمه‌ای است.

کلیدواژه‌ها: کشف تقلب، درمان تکمیلی، بیمه، شبکه عصبی، داده کاوی، شبکه عصبی پرسپترون، شبکه عصبی باور عمیق، شبکه عصبی بازگشتی

استناد به این مقاله: اسماعیلی، معصومه و ملکی نیا، محمد و پورابراهیمی، علیرضا . (۱۴۰۴). تشخیص تقلب در بیمه درمان تکمیلی با استفاده از الگوریتم‌های شبکه عصبی.مدیریت استراتژیک هوشمند, ۴(۳) , ۶۸-۴۷.

* نویسنده مسئول: M_malekinia@azad.ac.ir

مقدمه

در سال‌های اخیر، صنعت بیمه سلامت با افزایش چشم‌گیر فعالیت‌های متقلبانه مواجه شده است، به‌ویژه در حوزه بیمه‌های درمان تکمیلی که به دلیل ماهیت پیچیده و حجم بالای تراکنش‌ها، زمینه‌ای مساعد برای بروز تقلب فراهم می‌سازند. وقوع تقلب در بیمه نه تنها منجر به خسارت‌های مالی مستقیم برای شرکت‌های بیمه‌گر می‌شود، بلکه باعث افزایش نرخ حق‌بیمه برای بیمه‌گذاران و کاهش اعتماد عمومی به نظام سلامت نیز خواهد شد. از این‌رو، شناسایی و پیشگیری از تقلب‌های بیمه‌ای به یکی از چالش‌های اساسی برای بیمه‌گران، سیاست‌گذاران و پژوهشگران علوم داده تبدیل شده است. بیمه درمان تکمیلی که با هدف پوشش هزینه‌های خارج از تعهد بیمه‌های پایه طراحی شده است، به‌دلیل پیچیدگی فرایندهای رسیدگی، مشارکت واسطه‌هایی نظیر شرکت‌های ثالث رسیدگی‌کننده (TPA)، و همچنین نبود سامانه‌های یکپارچه اطلاعات سلامت و اشتراک‌گذاری داده، بیش از سایر شاخه‌های بیمه در معرض تقلب قرار دارد. در بسیاری از کشورها، به‌ویژه کشورهای در حال توسعه مانند ایران، نبود زیرساخت‌های مناسب برای ثبت الکترونیکی اطلاعات درمانی و فقدان سازوکارهای نظارت هوشمند، شناسایی تقلب را به کاری دشوار و پرهزینه تبدیل کرده است. روش‌های سنتی شناسایی تقلب، همچون ممیزی‌های دستی یا سیستم‌های مبتنی بر قواعد ثابت، اغلب در شناسایی الگوهای پیچیده، پنهان یا نوظهور ناکارآمد هستند. این روش‌ها معمولاً متکی بر دانش خبرگان و تجارب گذشته‌اند که در مواجهه با شیوه‌های نوین تقلب کارایی خود را از دست می‌دهند. علاوه بر این، چنین روش‌هایی معمولاً رویکردی واکنشی دارند و در محیط‌های حجیم و پویای داده توان مقیاس‌پذیری محدودی از خود نشان می‌دهند.

پیشرفت‌های اخیر در حوزه یادگیری ماشین، به‌ویژه یادگیری عمیق، چشم‌اندازهای جدیدی در حوزه شناسایی تقلب گشوده‌اند. مدل‌های یادگیری عمیق با توانایی استخراج خودکار الگوهای پیچیده و غیرخطی از داده‌های بزرگ و چندبعدی، در تشخیص ناهنجاری‌ها و طبقه‌بندی داده‌های پیچیده عملکرد قابل توجهی داشته‌اند. شبکه‌های عصبی پرسپترون چندلایه (MLP)، باور عمیق (DBN) و بازگشتی (LSTM) از جمله معماری‌هایی هستند که در مسائل مشابه مانند تقلب در بانکداری، تجارت الکترونیک و مخابرات با موفقیت استفاده شده‌اند. با وجود رشد فزاینده پژوهش‌ها در حوزه شناسایی تقلب با استفاده از یادگیری ماشین، تعداد مطالعات متمرکز بر بیمه درمان تکمیلی به‌ویژه در شرایط واقعی

کشورهای در حال توسعه، همچنان اندک است. کمبود داده‌های استاندارد، محدودیت‌های فنی و سازمانی و نبود زیرساخت‌های هوشمند از جمله موانعی هستند که پیاده‌سازی راهکارهای پیشرفته تحلیل داده را در این کشورها با دشواری همراه کرده‌اند. این پژوهش با هدف طراحی یک چارچوب داده‌محور برای شناسایی تقلب در بیمه درمان تکمیلی، به مقایسه عملکرد الگوریتم‌های یادگیری عمیق MLP، DBN و LSTM در شناسایی تقلب‌های احتمالی در داده‌های واقعی بیمه می‌پردازد. نتایج این مطالعه می‌تواند علاوه بر کمک به توسعه ادبیات نظری حوزه تحلیل تقلب در بیمه، راهکارهایی عملی برای شرکت‌های بیمه‌گر به منظور بهبود فرایندهای مدیریت تقلب و کاهش زیان‌های مالی ارائه دهد. در سال‌های اخیر استفاده از تحلیل داده و شبکه عصبی به عنوان یکی از روش‌های موثر در کشف تقلب در بیمه‌های مختلف از جمله شناسایی و کشف تقلب در بیمه‌نامه خودروها یا بیمه‌های درمانی مطرح شده است اما تاکنون از این الگوریتم برای بیمه درمان تکمیلی بهره گرفته نشده است. لذا مسأله اصلی پژوهش حاضر، ارائه رویکرد جامعی برای شناسایی تقلب در بیمه درمان تکمیلی است که شامل تحلیل داده درمانی و شخصی فرد و استفاده از شبکه عصبی برای شناسایی الگوهای تقلب باشد زیرا این رویکرد میتواند به مدیران بیمه و صاحبان سیاستهای حوزه بیمه درمان تکمیلی کمک کند تا به دقت بیشتری در شناسایی تقلب بپردازند و برنامه‌های خود را به طور موثرتری برای مقابله با تقلب‌ها طراحی کنند؛ دلیل این امر این است که روش‌های کشف تقلب باید دقیق، قابل اطمینان و قابل اعتماد باشند تا از آسیب به حقوق و منافع بیمه گران و شرکت‌های بیمه جلوگیری شود و به دلیل پیچیدگی و حجم بالای داده‌های مربوط به بیمه درمان تکمیلی، استفاده از روش‌های پیشرفته مانند شبکه‌های عصبی برای کشف تقلب در بیمه درمان تکمیلی می تواند راهکاری بسیار موثر باشد. از سوی دیگر برای ارزیابی دقت و عملکرد مدل، از داده‌های واقعی بیماران استفاده شده است تا بتوان یک ارزیابی دقیق و شفاف از مدل داده درمانی فرد ارائه شود. از سوی دیگر در مدل ارائه شده، به توانایی شناسایی الگوهای تقلب در بیمه درمان تکمیلی توجه شده است. در این پژوهش، شناسایی الگوهای مشترک تقلب در بیمه درمان تکمیلی و کمک به مدیران بیمه درمان تکمیلی برای طراحی برنامه‌های کنترل تقلب است تا منجر به بهبود تجربه مشتریان از خدمات بیمه از طریق تصمیم گیری‌های هوشمندانه بر اساس مدل داده و بررسی نحوه تاثیر عوامل مختلف نظیر سن، سابقه بیماری، ویژگی‌های زندگی بر ریسک و کشف تقلب بیمه درمان است.

پیشینه تحقیق

شناسایی تقلب در بیمه یکی از حوزه‌های پژوهشی مهم و پرچالش در علوم داده و یادگیری ماشین است که طی سال‌های اخیر توجه گسترده‌ای را به خود جلب کرده است. در این بخش، به بررسی پژوهش‌های کلیدی انجام شده در زمینه شناسایی تقلب بیمه‌ای و به ویژه کاربرد الگوریتم‌های یادگیری عمیق پرداخته می‌شود.

چنگ و همکاران[1](۲۰۲۴) به تحقیق در مورد مدل ضد کلاهبرداری بیمه پزشکی بر اساس نمودار دانش و ترکیب الگوریتم‌های Louvain و NetworkX پرداخته‌اند. در این مقاله، یک مدل بدون نظارت بر اساس نمودار دانش و الگوریتم Louvain برای شناسایی کلاهبرداری باند بیمه پزشکی پیشنهاد شده است. مدل پیشنهادی تحقیق، در مرحله اول، یک نمودار دانش بر اساس سوابق پزشکی با استفاده از الگوریتم NetworkX برای ایجاد یک نمودار دانش از تقلب ضد باند در بیمه پزشکی می‌سازد که خلاصه کردن قوانین ریسک را پس از تقسیم جامعه تسهیل می‌کند. سپس، الگوریتم Louvain به شبکه ارتباط بیمار و پزشک برای کشف جوامع اعمال می‌شود و سپس کل نمودار دانش را به ترتیب به چهار سطح جامعه با ریسک بالا، متوسط، پایین و بدون خطر آشکار تقسیم می‌کند. برای نشان دادن برتری مدل پیشنهادی، نتایج مدل با سایر مدل‌های بدون نظارت بر روی مجموعه داده‌های متعدد برای شناسایی کلاهبرداری باند مقایسه شده است. طبق نتایج تحقیق، با مقایسه نرخ پارتیشن صحیح، برتری مدل پیشنهادی در تحقیق شناسایی باندهای کلاهبرداری بیمه‌های پزشکی نشان داده شده است.

هونگ و همکاران[2](۲۰۲۴) به تشخیص تقلب بیمه سلامت بر اساس یادگیری ساختار گراف ناهمگن چند کانالی (MHGSL) پرداخته‌اند. روش MHGSL نموداری از داده‌های بیمه سلامت از نهادهای مختلف مانند بیماران، بخش‌ها و داروها می‌سازد و از یادگیری ساختار نمودار برای استخراج ساختار توپولوژیکی، ویژگی‌ها و اطلاعات معنایی برای ساخت نمودارهای متعددی استفاده می‌کند که تنوع و پیچیدگی داده‌ها را منعکس می‌کند. در این تحقیق از روش‌های یادگیری عمیق مانند شبکه‌های عصبی نمودار ناهمگن و شبکه‌های عصبی کانولوشن گراف برای ترکیب انتقال اطلاعات چند کانالی و ترکیب ویژگی‌ها برای تشخیص ناهنجاری‌ها در داده‌های بیمه سلامت استفاده شده است. نتایج نشان داد روش MHGSL به دقت بالایی در تشخیص تقلب بالقوه دست می‌یابد که بهتر از روش‌های

[1] Cheng et al.
[2] Hong et al.

موجود است و قادر است به سرعت و با دقت بیماران دارای رفتارهای متقلبانه را شناسایی کند تا از دست دادن وجوه بیمه سلامت جلوگیری شود.

با گسترش یادگیری ماشین، مدل‌های نظارت شده و غیرنظارت شده مانند درخت تصمیم، ماشین بردار پشتیبان (SVM)، و خوشه‌بندی برای تشخیص الگوهای غیرعادی در داده‌های بیمه مورد استفاده قرار گرفتند. پژوهش‌ها نشان داده‌اند که این روش‌ها نسبت به رویکردهای سنتی دقت و قابلیت تشخیص بهتری دارند، اما همچنان محدودیت‌هایی از قبیل نیاز به مهندسی ویژگی پیچیده و عدم توانایی در استخراج روابط غیرخطی چندبعدی دارند (Chen & Li, 2017).

در سال‌های اخیر، الگوریتم‌های یادگیری عمیق به دلیل قابلیت استخراج خودکار ویژگی‌ها و مدل‌سازی پیچیدگی‌های داده‌های حجیم و چندبعدی، به عنوان رویکردی نوین در شناسایی تقلب مطرح شده‌اند. مطالعه Wang و همکاران (۲۰۱۹) با استفاده از شبکه‌های عصبی پرسپترون چندلایه (MLP) در داده‌های بیمه سلامت، نشان داد که این مدل‌ها قادرند الگوهای پنهان تقلب را با دقت بالاتری نسبت به مدل‌های کلاسیک شناسایی کنند.

همچنین، باورهای عمیق (Deep Belief Networks - DBN) به عنوان یکی از معماری‌های پیشرفته یادگیری عمیق، در پژوهش Li و Zhao (2020) برای کشف تقلب در بیمه اتومبیل به کار رفتند و نتایج بهبود چشمگیری در کاهش نرخ تشخیص خطا نشان دادند. این مدل‌ها به دلیل ساختار لایه‌ای خود توانایی مدل‌سازی توزیع‌های پیچیده داده‌ها را دارند که برای مسائل تشخیص تقلب بسیار حیاتی است.

علاوه بر آن، شبکه‌های عصبی بازگشتی (Recurrent Neural Networks - RNN) به دلیل قابلیت پردازش داده‌های توالی‌دار و زمانی، در پژوهش‌های مربوط به تشخیص تقلب در تراکنش‌های مالی و بیمه کاربرد یافته‌اند. پژوهش Zhang و همکاران (۲۰۲۱) با بهره‌گیری از معماری LSTM که یکی از انواع RNN است، موفق به بهبود دقت تشخیص تقلب در داده‌های بیمه درمانی شدند و نشان دادند که تحلیل توالی زمانی داده‌ها می‌تواند به شناسایی بهتر الگوهای غیرعادی کمک کند.

با این وجود، بیشتر تحقیقات انجام شده بر روی داده‌های عمومی و در برخی موارد داده‌های شبیه‌سازی شده است و پژوهش‌های کاربردی و مبتنی بر داده‌های واقعی بیمه درمان تکمیلی، خصوصاً در کشورهای در حال توسعه، اندک و محدود به چند مطالعه بوده است (Rahimi et al., 2022). همچنین، چالش‌هایی نظیر عدم دسترسی به داده‌های کامل، ناهمگنی

داده‌ها و کمبود زیرساخت‌های تحلیل داده‌های بزرگ، از موانع مهم در توسعه این حوزه به شمار می‌روند. در ایران نیز مطالعات محدودی در زمینه شناسایی تقلب بیمه‌ای انجام شده است که عمدتاً بر روش‌های آماری و یادگیری ماشین کلاسیک متمرکز بوده‌اند و کمتر به استفاده از تکنیک‌های یادگیری عمیق توجه شده است

احمدی (۱۴۰۲) به ارائه الگوریتمی مبتنی بر گراف برای کشف تقلب در بیمه درمان پرداخته است. این پژوهش به بررسی استفاده از نظریه گراف برای کشف تقلب در صنعت بیمه ی درمان می‌پردازد. با استخراج اطلاعات از پایگاه داده‌های مربوط و ساخت گراف شبکه‌ی کلی، الگوهای موجود در آن تحلیل شده و موارد مشکوک تقلبی شناسایی می گردند. روش پیشنهادی بر روی داده‌های واقعی اجرا و نتایج نشان داد که این روش قادر به کشف تقلب‌ها با ۹۵ درصد دقت بوده که نسبت به روش‌های موجود بهبود یافته است.

غفاری (۱۴۰۲) به معرفی و اهمیت کاربرد الگوریتم یادگیری ماشین در تشخیص تقلب بیمه درمان تکمیلی پرداخته است. هدف اصلی این مطالعه معرفی کاربرد الگوریتم یادگیری ماشین برای کشف تقلب در قراردادهای بیمه بوده است. از آنجایی که انواع عوامل کلاهبرداری در صنعت بیمه سلامت متفاوت است، در این مطالعه از دو روش Auto Encoder و Auto Encoder Variational برای شناسایی عناصر اصلی تقلب و رتبه بندی متغیر تأثیرگذار در این موضوع از قراردادهای بیمه استفاده شده است. نتایج تحقیق نشان داد الگوریتم یادگیری ماشین میتواند عملکرد موثری در تشخیص تقلب بیمه درمان تکمیلی داشته باشد و در تحقیقات آتی می‌توان از این روش‌ها برای تشخیص تقلب بیمه درمان تکمیلی استفاده نمود.

رحیم خانی و منطقی پور (۱۴۰۱) در پژوهش خود با انتخاب روش نظام مندی که توسط سیلوا معرفی شده به مرور نظام مند الگوریتم داده کاوی برای کشف تقلب در بیمه درمانی پرداخته اند که نتایج به شرح ذیل بوده است: تشخیص تقلب در مراقبت‌های بهداشتی عمدتا با استفاده از روش‌های یادگیری ماشین و داده کاوی بوده است. از الگوریتم‌های تحت نظارت، الگوریتم‌های بدون نظارت و الگوریتم‌های ترکیبی در پژوهش‌ها استفاده شده است. بیشتر مطالعات تا سال ۲۰۱۷ از الگوریتم‌های بدون نظارت استفاده شده است. مطالعات سال ۲۰۱۷ به بعد از الگوریتم تحت نظارت استفاده شده است. روش‌های نزدیک ترین همسایه، شبکه عصبی و روشهای مبتنی بر قانون هم برای الگوریتم‌های تحت نظارت و هم برای الگوریتم‌های بدون نظارت در کشف تقلب در بیمه درمانی استفاده شده است. در بسیاری از

الگوریتم‌های ترکیبی از روشهای شبکه عصبی استفاده شده اما هیچ الگو یا روش استانداردی برای پوشش همه موارد وجود ندارد.

روش شناسی پژوهش

هدف اصلی این پژوهش طراحی مدلی برای شناسایی تقلب در بیمه درمان تکمیلی با بهره‌گیری از الگوریتم‌های یادگیری ماشین است. رویکرد پژوهش بر پایه داده‌کاوی و تحلیل داده‌های واقعی بیمه‌ای با استفاده از شبکه‌های عصبی پیشرفته است. در این بخش مراحل انجام پژوهش، ساختار مدل‌ها و شیوه ارزیابی عملکرد آن‌ها به‌تفصیل شرح داده می‌شود.

۱. جمع‌آوری و مشخصات داده‌ها

داده‌های پژوهش از سامانه اطلاعات خسارات بیمه درمان تکمیلی یک شرکت بیمه در ایران استخراج شده‌اند. این داده‌ها شامل ۲۰۰۰۰ رکورد مربوط به درخواست‌های پرداخت خسارت طی یک بازه یکساله هستند.

ویژگی‌های کلیدی داده‌ها شامل:

اطلاعات فردی بیمه‌شده: سن، جنسیت، وضعیت تأهل، نوع قرارداد بیمه

ویژگی‌های درخواست خسارت: نوع خدمت درمانی (آزمایش، دارو، بستری، ویزیت، تصویربرداری و غیره)، مبلغ درخواستی، تعداد مراجعات، نوع مرکز درمانی (دولتی/خصوصی)، زمان ارسال درخواست

نمونه‌های شامل تقلب جزء یکی از موارد زیر هستند:

- دریافت خسارت در روز شروع بیمه نامه: اعلام خسارت در روزهای ابتدای بیمه نامه احتمال کتمان وضعیت جسمانی در هنگام عقد قرارداد یا تغییر جزئی تاریخ خسارت را بالا می برد
- تعداد خسارت زیاد طی چند روز متوالی: وقوع خسارت بسیار زیاد در طی چند روز نشان دهنده رفتاری نامتعارف در بانک اطلاعاتی بوده و نشان دهنده احتمال بالای بروز تقلب است
- وجود خسارت در دو یا چند شهر در یک روز: معمولا احتمال اینکه یک بیمه گذار در یک روز در دو یا چند شهر اعلام خسارت نماید احتمال بالای تقلب را نشان می دهد
- وقوع خسارت در روز آخر بیمه نامه: در این نوع از خسارت‌ها مانند نوع شماره یک احتمال وجود تقلب بیشتر است

برچسب هدف (Label): تقلبی (۱) یا غیرتقلبی (۰)

مجموعه داده ها

داده‌های مورد استفاده شامل پنج مجموعه داده به شرح ذیل بوده است:

- بیمه نامه: اطلاعات بیمه نامه‌های صادر شده
- بیمه شدگان: اطلاعات بیمه شده‌های مجموعه بیمه نامه
- خسارات: جزئیات خسارت‌های پرداختی هر بیمه نامه
- بیماری ها: انواع بیماری‌هایی که برای آنها خسارت پرداخت شده
- شعب بیمه: اطلاعات شعبه‌های صادر کننده بیمه نامه و حواله خسارت

جدول ۱: نمونه ای از داده های ورودی (بیمه شدگان)

2- ID	سن بیمه شده	مبلغ حواله	تقلبی (برچسب)
۱	۴۵	۸۵۰,۰۰۰	۰
۲	۶۰	۱,۲۰۰,۰۰۰	۱
۳	۳۰	۷۰۰,۰۰۰	۰
...	...	...	...

با استفاده از کلیدهای خارجی مجموعه‌ها و در نرم افزار SQL Server داده‌ها در هم ادغام شده و یک مجموعه ترکیبی جدید از تمام مجموعه‌ها ایجاد شد.

۲. پیش‌پردازش داده‌ها

مراحل پیش‌پردازش به‌منظور آماده‌سازی داده‌ها برای آموزش مدل‌ها شامل مراحل زیر بود:

حذف داده‌های ناقص و پرت: درخواست‌هایی با اطلاعات ناقص یا غیرمنطقی (مانند مبلغ منفی) حذف شدند.

نرمال‌سازی داده‌های عددی: از روش Min-Max Scaling برای متغیرهای پیوسته مانند سن و مبلغ استفاده شد تا مقادیر بین صفر و یک قرار گیرند. رمزگذاری متغیرهای طبقه‌ای: متغیرهایی مانند نوع خدمت، جنسیت و نوع مرکز به‌صورت One-Hot Encoding تبدیل شدند.

تقسیم داده‌ها: داده‌ها به‌صورت تصادفی به دو بخش تقسیم شدند:

داده‌های آموزش: ٪۸۰

داده‌های آزمون: ٪۲۰

استخراج ایندکس‌های افراد متقلب: با استفاده از پیش‌بینی‌های انجام‌شده، ایندکس‌های بیمه‌شدگان متقلب استخراج می‌شوند.

شناسایی سن و مبلغ واقعی: اطلاعات مربوط به سن و مبلغ حواله واقعی افراد متقلب از داده‌های اصلی استخراج و نمایش داده می‌شود.

داده‌های دنیای واقعی ناقص، ناسازگار و دارای نویز هستند. روال‌های پاک کردن داده‌ها یا پالایش داده‌ها سعی در پر کردن مقادیر مفقوده، حذف و یا نرم کردن نویز و اصلاح ناسازگاری در داده‌ها دارند. کاهش ابعاد نیز یکی از مراحل پیش پردازش و در هنگام استخراج ویژگی است. دلیل کاهش ابعاد را می توان راحت‌تر شدن تحلیل‌های بعدی، افزایش عملکرد جدا کننده بر اساس نمایش بهتر (پایدارتر)، حذف اطلاعات تکراری یا غیر مرتبط و یا تلاشی برای کشف ساختار اساسی با به دست آوردن نمایش گرافیکی از داده‌ها دانست.

شکل ۱: دیاگرام مدل پیشنهادی تحقیق

۳. طراحی مدل‌های یادگیری عمیق

سه نوع شبکه عصبی برای مدل‌سازی انتخاب شدند که هر یک قابلیت خاصی برای کشف الگوهای پیچیده و شناسایی رفتارهای مشکوک دارند:

شبکه عصبی پرسپترون چندلایه (MLP)

ساختار: ورودی با تعداد نورون برابر ویژگی‌ها، ۳ لایه پنهان با ۶۴، ۳۲، و ۱۶ نورون

تابع فعال‌سازی: ReLU در لایه‌های میانی، Softmax در خروجی

بهینه‌ساز: Adam

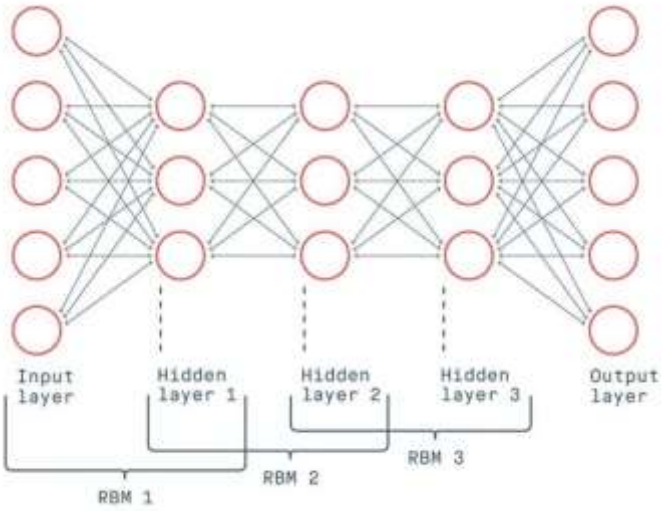
تابع خطا: Binary Cross-Entropy

شبکه باور عمیق (DBN)

شامل چندین RBM به‌صورت پشته‌ای برای استخراج ویژگی‌ها

فاز پیش‌آموزش غیربرخط (unsupervised) با Contrastive Divergence

طبقه‌بندی نهایی با لایه Logistic Regression



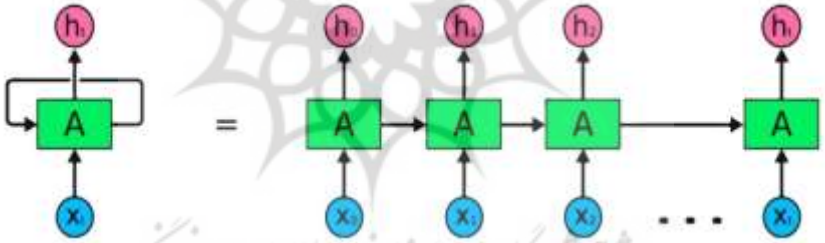
شکل ۲: توانایی مدل‌سازی روابط پنهان پیچیده در داده‌ها

شبکه عصبی بازگشتی (RNN با LSTM)

مناسب برای درک وابستگی‌های زمانی بین مراجعات پی‌درپی

شامل یک لایه LSTM با ۶۴ واحد حافظه و Dropout

امکان تحلیل دنباله‌ای مراجعات درمانی و الگوهای زمانی مشکوک



شکل ۳: شبکه عصبی بازگشتی

۴. معیارهای ارزیابی

برای مقایسه عملکرد مدل‌ها از چهار معیار استاندارد یادگیری ماشین استفاده شد:

دقت (Accuracy): نسبت کل پیش‌بینی‌های درست

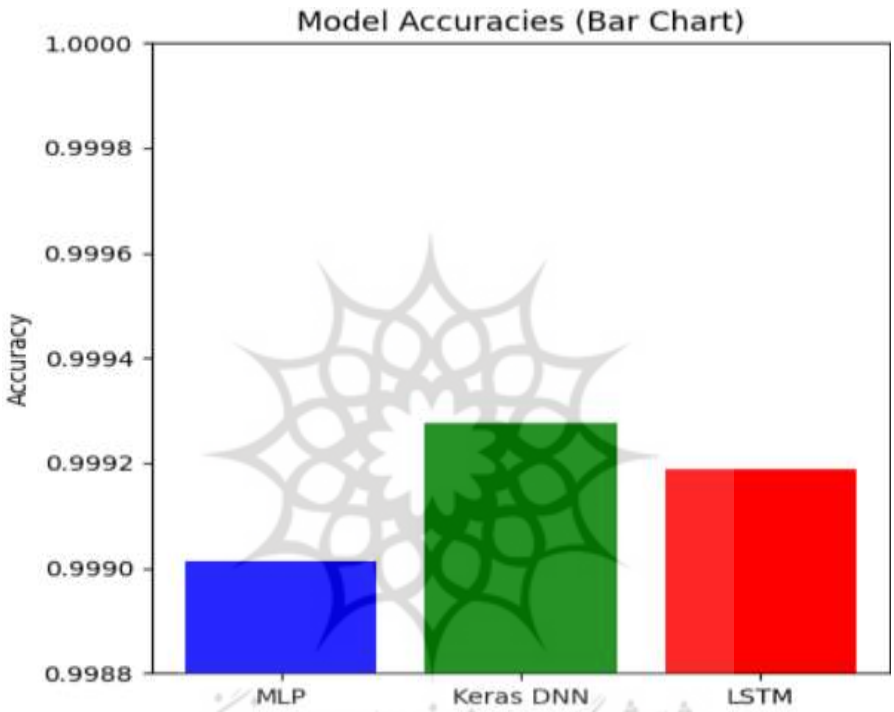
دقت مثبت (Precision): نسبت پیش‌بینی‌های درست از بین همه مواردی که به‌عنوان تقلبی شناسایی شده‌اند

بازیابی (Recall): نسبت پیش‌بینی‌های درست تقلب از بین همه موارد تقلبی واقعی

امتیاز F1 (F1-Score): میانگین موزون Precision و Recall

مدل	Precision	Recall	F1-Score	Accuracy
MLP	0.90	0.91	0.91	89.0
Keras DNN	0.93	0.89	0.93	85.3
LSTM	0.92	0.93	0.95	92.1

نتایج مقایسه دقت سه مدل



شکل ۴: نتایج مقایسه دقت سه مدل

جمع‌بندی روش‌شناسی

مدل RNN به دلیل ساختار حافظه‌دار و توانایی شناسایی وابستگی‌های زمانی، در شناسایی رفتارهای مشکوک و الگوهای تقلب عملکرد بهتری نسبت به سایر مدل‌ها داشت. با توجه به پیچیدگی رفتار تقلب در بیمه، استفاده از معماری‌های پیشرفته یادگیری عمیق می‌تواند نقش مؤثری در بهبود دقت شناسایی تقلب و کاهش زیان شرکت‌های بیمه ایفا کند. هر یک از این مدل‌ها می‌توانند در موقعیت‌های مختلف، مزایای خاص خود را به‌دست آورده و به

افزایش دقت و کارایی سیستم‌های شناسایی تقلب کمک کنند. این تحلیل جامع و متنوع، نه تنها به شناسایی تقلب‌های آشکار کمک می‌کند، بلکه می‌تواند به‌عنوان ابزاری موثر برای مقابله با تقلب‌های پیچیده و زمان‌بر در صنعت بیمه عمل نماید. "نتایج این مطالعه نشان داد که انتخاب مدل مناسب برای هر نوع داده از مسائل بسیار با اهمیت است. در اینجا، مدل LSTM در پردازش داده‌های ترتیبی، بهترین عملکرد را در مقایسه با مدل‌های DNN و DBN ارائه داد. با این حال، برای داده‌هایی که وابستگی‌های زمانی ندارند، مدل DNN می‌تواند گزینه بهتری باشد. همچنین، مدل DBN که به‌ویژه در استخراج ویژگی‌های پیچیده موفق عمل کرده است، در برخی از مشکلات دیگر نیز نتایج مثبتی را به‌دست آورد. در نهایت، توصیه می‌شود که برای مسائل مختلف، علاوه بر انتخاب مدل مناسب، از تکنیک‌های تنظیم پارامترها و بهینه‌سازی مدل‌ها برای بهبود عملکرد استفاده شود." همچنین نتایج به دست آمده شامل مدل Keras DNN با دقت ۰,۹۹۹۳ ، در حالی که MLP و LSTM به ترتیب دقت‌های ۰,۹۹۹۰ و ۰,۹۹۹۲ را به دست آوردند. این نتایج نشان می‌دهند که تمامی مدل‌ها قادر به شناسایی دقیق تقلب در داده‌های بیمه درمان تکمیلی هستند و تفاوت‌های جزئی در عملکرد آن‌ها ممکن است به عوامل مختلفی مانند نوع ویژگی‌ها یا ساختار داده‌ها وابسته باشد. با توجه به نتایج، می‌توان نتیجه گرفت که الگوریتم‌های شبکه عصبی می‌توانند ابزارهای بسیار مؤثری برای شناسایی تقلب در بیمه درمان تکمیلی باشند. این مدل‌ها قادرند الگوهای پیچیده و غیرخطی موجود در داده‌های بیمه را شناسایی کرده و از این رو، می‌توانند به طور چشمگیری در کاهش تقلب‌های بیمه‌ای و بهبود فرآیندهای نظارتی مفید واقع شوند. برای بهبود بیشتر عملکرد این مدل‌ها، استفاده از تکنیک‌های بهینه‌سازی و تنظیمات دقیق‌تر پارامترها می‌تواند کمک شایانی به افزایش دقت و قابلیت اعتماد مدل‌ها داشته باشد.

جدول ۲: پرسپترون

پرسپترون	مناسب برای روابط ساده و خطی
DNN	برای روابط پیچیده‌تر و غیرخطی
LSTM	برای تحلیل داده‌های توالی‌دار و تاریخچه‌ای

یافته پژوهش

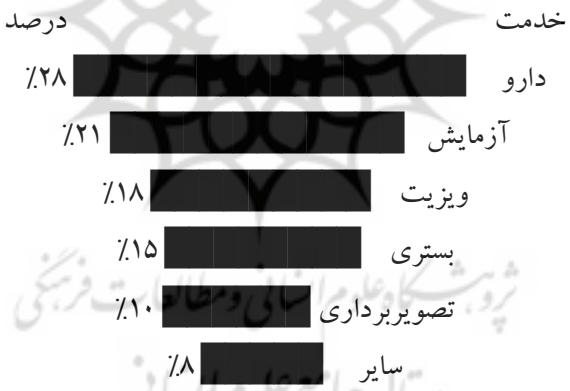
۱. مشخصات آماری داده‌ها

داده‌های مورد استفاده شامل ۲۰۰۰۰ رکورد مربوط به درخواست‌های خسارت بیمه درمان تکمیلی طی یک بازه یکساله بوده است. توزیع برخی متغیرهای کلیدی به شرح زیر است:

جدول۳: توزیع برخی متغیرهای کلیدی

متغیر	میانگین / درصد	توضیحات
سن بیمه‌شده	۴۲٫۸ سال	دامنه: ۱ تا ۸۷ سال
جنسیت بیمه‌شده	٪۵۳ زن	٪۴۷ مرد
وضعیت تأهل	٪۶۵ متأهل	٪۳۵ مجرد
نوع خدمت درمانی	٪۲۸ دارو، ٪۲۱ آزمایش، ٪۱۸ ویزیت، ٪۱۵ بستری، ٪۱۰ تصویربرداری، ٪۸ سایر	
نوع مرکز درمانی	٪۶۱ خصوصی،	٪۳۹ دولتی
برچسب تقلبی	٪۷٫۴ موارد تقلبی	حدود ۳٬۸۰۰ مورد تقلبی ثبت‌شده

نمودار زیر توزیع نوع خدمات درمانی را نشان می‌دهد:



نمودار۱: توزیع نوع خدمات درمانی

۲. عملکرد مدل‌های یادگیری

برای ارزیابی عملکرد سه مدل یادگیری (MLP، DBN و RNN) از چهار معیار رایج استفاده شد. نتایج ارزیابی روی داده‌های آزمون در جدول زیر آمده است:

نتایج به دست آمده از آزمایش مدل‌ها نشان‌دهنده عملکرد بسیار بالا و مشابه این مدل‌ها در شناسایی تقلب بوده است. مدل Keras DNN بهترین دقت را با مقدار ۰,۹۹۹۳ به دست آورد و پس از آن، مدل‌های MLP با دقت ۰,۹۹۹۰ و LSTM با دقت ۰,۹۹۹۲ قرار گرفتند. این دقت‌های بالا بیانگر این است که تمامی مدل‌ها توانسته‌اند به خوبی ویژگی‌های پیچیده و غیرخطی موجود در داده‌های بیمه درمان تکمیلی را شناسایی کرده و تقلب‌ها را با دقت بالایی تشخیص دهند. با این حال، تفاوت‌های جزئی در عملکرد مدل‌ها ممکن است به علت تفاوت در ساختار داده‌ها یا ویژگی‌های خاص هر مدل باشد. تحلیل حساسیت مدل‌ها به ورودی‌ها می‌تواند به ارزیابی این موضوع بپردازد که چطور تغییرات در داده‌های ورودی و ویژگی‌ها بر دقت و عملکرد مدل تاثیر می‌گذارند. می‌توان برای این منظور از **روش‌های حساسیت** مانند **Shapley Values** استفاده کرد تا درک دقیقی از اهمیت ویژگی‌های مختلف داده‌ها در فرآیند شناسایی تقلب به دست آورید. نتایج به دست آمده نشان‌دهنده دقت بالای مدل‌های شبکه عصبی در شناسایی تقلب در بیمه درمان تکمیلی است. به‌ویژه مدل Keras DNN با دقت ۰,۹۹۹۳، عملکرد بهتری نسبت به سایر مدل‌ها نشان داد. این عملکرد بالا نشان‌دهنده توانایی مدل‌ها در شناسایی ویژگی‌های پیچیده و غیرخطی موجود در داده‌های بیمه درمان تکمیلی است که تشخیص آن‌ها برای الگوریتم‌های سنتی دشوار است. مدل‌های MLP و LSTM نیز عملکرد بسیار خوبی داشتند و دقت بالایی را در شناسایی تقلب‌ها نشان دادند. این مدل‌ها می‌توانند در سیستم‌های واقعی و در زمان‌های کوتاه به طور مؤثر به کار گرفته شوند و از آنجا که دقت بالایی دارند، می‌توانند به طور جدی در کاهش تقلب‌ها در بیمه‌های درمان تکمیلی مفید باشند. همچنین مدل LSTM در پردازش داده های ترتیبی عملکرد بهتری نشان داده است.

۳. تحلیل دقیق‌تر عملکرد مدل‌ها

MLP علی‌رغم سرعت آموزش و سادگی پیاده‌سازی، به‌دلیل نداشتن ساختار حافظه‌دار، در شناسایی الگوهای زمانی ضعیف‌تر عمل کرد.

DBN در استخراج ویژگی‌های پنهان غیرخطی عملکرد خوبی داشت و به دلیل آموزش لایه‌به‌لایه، توانست روابط پنهان بین متغیرها را بهتر درک کند.

RNN با LSTM، بهترین عملکرد را ارائه داد. این مدل با قابلیت درک دنباله‌ای و زمانی مراجعات پزشکی، توانست رفتارهای مشکوک متوالی را به‌خوبی شناسایی کند.

۴. تحلیل خطای مدل‌ها

برای بررسی جزئی‌تر، نرخ خطای مثبت کاذب (False Positive Rate) و منفی کاذب (False Negative Rate) نیز استخراج شد:

جدول ۳: نرخ خطای مثبت کاذب و منفی کاذب

مدل	نرخ مثبت کاذب (FPR)	نرخ منفی کاذب (FNR)
MLP	٪۶٫۸	٪۱۳
DBN	٪۴٫۲	٪۷
RNN	٪۳٫۱	٪۶

جمع‌بندی یافته‌ها

نتایج نشان می‌دهد که استفاده از مدل‌های یادگیری عمیق می‌تواند به‌صورت قابل توجهی دقت شناسایی تقلب در بیمه درمان تکمیلی را افزایش دهد. در این پژوهش:

مدل RNN با LSTM به‌عنوان دقیق‌ترین مدل معرفی شد و برای کاربردهای واقعی در سیستم‌های ضدتقلب بیمه‌ای پیشنهاد می‌شود.

معماری‌های عمیق مانند DBN و RNN با توانایی کشف الگوهای پنهان و وابستگی‌های زمانی، در مقایسه با مدل‌های سنتی یا ساده‌تر مانند MLP، عملکرد قابل‌توجهی داشتند

نتیجه‌گیری

۱. تحلیل نتایج مدل‌سازی

هدف این پژوهش طراحی یک مدل هوشمند برای شناسایی تقلب در بیمه درمان تکمیلی با استفاده از الگوریتم‌های یادگیری ماشین بود. سه مدل پرکاربرد شامل شبکه پرسپترون چندلایه (MLP)، شبکه باور عمیق (DBN)، و شبکه عصبی بازگشتی (RNN) بر روی داده‌های واقعی بیمه‌ای اجرا و ارزیابی شدند. تحلیل عملکرد این مدل‌ها نشان‌دهنده مزایا و محدودیت‌های هر یک در شناسایی رفتارهای مشکوک بود.

مدل RNN (LSTM) با ساختار حافظه‌دار خود، توانسته است الگوهای زمانی در مراجعات بیمه‌ای را شناسایی کرده و رفتارهای غیرعادی (نظیر تکرار الگوی خاص مراجعه یا درخواست هزینه با فاصله زمانی مشکوک) را بهتر از سایر مدل‌ها تشخیص دهد.

مدل DBN به دلیل بهره‌گیری از چند لایه RBM در فاز پیش‌آموزش، قدرت بالایی در استخراج ویژگی‌های پنهان و همبستگی‌های پیچیده بین متغیرها دارد.

مدل MLP علی‌رغم ساختار ساده‌تر و سرعت بالاتر در آموزش، در مواجهه با داده‌های پیچیده‌تر و دارای وابستگی زمانی، دقت کمتری دارد.

۲. تحلیل کاربردی و صنعتی نتایج

استفاده از مدل‌های یادگیری عمیق در صنعت بیمه مزایای فراوانی دارد. در این پژوهش نشان داده شد که این مدل‌ها می‌توانند:

الف. شناسایی دقیق‌تر تقلبات پیچیده

برای مثال، یک بیمار ممکن است طی دو ماه در چند مرکز درمانی مختلف مراجعه کرده باشد و در هر بار خدمات متفاوتی (دارو، ویزیت، آزمایش) دریافت کرده باشد. شناسایی چنین الگوهایی که در ظاهر طبیعی ولی در عمق داده‌ها مشکوک هستند، تنها با مدل‌هایی مانند RNN (LSTM)که توانایی تحلیل دنباله‌ای دارند، ممکن می‌شود.

ب. کاهش بار کارشناسان بررسی خسارت

در حال حاضر، بسیاری از کارشناسان بیمه مجبورند به‌صورت دستی اطلاعات مراجعات بیمه‌شدگان را بررسی و تحلیل کنند. مدل‌های هوشمند می‌توانند پیش‌پردازش اولیه را انجام داده و تنها موارد مشکوک را به کارشناس ارسال کنند، که باعث صرفه‌جویی در وقت و هزینه می‌شود.

ج. ارتقاء سیستم‌های تصمیم‌یار بیمه‌ای

با تلفیق خروجی مدل‌های هوشمند با سامانه‌های مدیریت خسارت، شرکت بیمه می‌تواند در لحظه تصمیم‌گیری‌های مناسبی اتخاذ کند، مثلاً جلوگیری از پرداخت یا ارجاع برای بررسی دقیق‌تر.

۳. ملاحظات و محدودیت‌ها

با وجود نتایج مثبت، پژوهش با چالش‌هایی نیز مواجه بوده است:

داده‌های ناقص و مغشوش: بخشی از داده‌ها دارای اطلاعات ناقص یا غیرواقعی بوده‌اند (مثلاً مقادیر منفی یا سن غیرمنطقی)، که نیازمند پاک‌سازی دقیق بودند.

عدم وجود ویژگی‌های اجتماعی-اقتصادی بیمه‌شده: عواملی چون شغل، تحصیلات یا وضعیت اقتصادی می‌توانند در تقلب موثر باشند اما در داده‌های موجود لحاظ نشده بودند.

عدم تعادل کلاس‌ها: تقلب در واقعیت بسیار کمتر از موارد سالم است (یعنی داده‌ها دچار عدم تعادل بودند)، که این مسأله با استفاده از تکنیک‌هایی مانند class weighting یا oversampling مدیریت شد.

۴. پیشنهادهایی برای پژوهش‌های آینده

توسعه مدل‌های یادگیری ترکیبی (Hybrid Deep Models) مانند ترکیب RNN با CNN برای شناسایی الگوهای متنی (مثلاً در نسخه پزشک یا فاکتور)

استفاده از Graph Neural Networks برای تحلیل روابط بین بیمه‌گذاران، پزشکان، داروخانه‌ها و مراکز درمانی

پیاده‌سازی سامانه هشدار سریع (Early Warning System) مبتنی بر یادگیری آنلاین (Online Learning)

اعتبارسنجی مدل‌ها در سایر شاخه‌های بیمه مانند بیمه خودرو یا بیمه عمر

استفاده از مدل‌های پیشرفته‌تر شبکه عصبی مانند Transformers و Autoencoders است که قادر به شناسایی الگوهای پیچیده‌تری در داده‌ها هستند. این مدل‌ها می‌توانند در شناسایی تقلب‌هایی که الگوریتم‌های سنتی قادر به تشخیص آن‌ها نیستند، کارآمدتر باشند.

بهینه‌سازی پارامترهای مدل با استفاده از تکنیک‌هایی مانند Grid Search یا Bayesian Optimization می‌توان پارامترهای مدل‌ها را بهینه‌سازی کرد تا عملکرد بهتری بدست آید.

برای بهبود عملکرد مدل‌ها، پیشنهاد می‌شود که از داده‌های بیشتر و متنوع‌تری استفاده شود. داده‌هایی که از منابع مختلف جمع‌آوری شده و شامل ویژگی‌های پیچیده‌تری باشند، می‌توانند به مدل‌ها کمک کنند تا به‌طور دقیق‌تری الگوهای تقلب را شناسایی کنند. داده‌های دارای ویژگی‌های غیرساختاری مانند متغیرهای محیطی و اجتماعی می‌توانند ویژگی‌های جدیدی برای مدل‌ها ایجاد کنند.

نمونه‌های کاربردی برای پیاده‌سازی

فرض کنید یک شرکت بیمه درمان تکمیلی با تعداد زیادی پرونده از بیمه‌گذاران دارد و قصد دارد تقلب‌های احتمالی را شناسایی کند. مدل‌های شبکه عصبی مورد بررسی در این تحقیق می‌توانند در این سناریو با تحلیل داده‌ها به شناسایی الگوهای مشکوک و افراد متقلب بپردازند. به‌عنوان مثال: اگر فردی با چندین بار ادعای درمان مشابه در مدت زمانی کوتاه چندین بار برای دریافت هزینه مراجعه کند، مدل‌ها می‌توانند این رفتار را به عنوان نشانه‌ای از تقلب شناسایی کنند. همچنین، در صورتی که فردی به دفعات از خدمات خاصی استفاده کند که به‌طور غیرمعمولی گران هستند، مدل‌های شناسایی تقلب می‌توانند این نوع ناهنجاری‌ها را شناسایی کنند.

جمع‌بندی نهایی

پژوهش حاضر با هدف ارائه روشی نوین برای شناسایی تقلب در بیمه درمان تکمیلی، سه مدل مبتنی بر یادگیری عمیق را با داده‌های واقعی مورد آزمایش قرار داد. نتایج به‌وضوح نشان داد که مدل RNN (LSTM) به‌دلیل قابلیت تحلیل ترتیبی و یادآوری اطلاعات گذشته، نسبت به سایر مدل‌ها عملکرد برتری دارد. این مدل قادر است الگوهای مشکوک و پیچیده را در میان داده‌های حجیم تشخیص دهد. به‌کارگیری الگوریتم‌های یادگیری عمیق می‌تواند نه‌تنها نرخ کشف تقلب را افزایش دهد، بلکه با کاهش هزینه‌های عملیاتی و بهبود کیفیت خدمات، سودآوری شرکت‌های بیمه را در بلندمدت تضمین کند. آینده صنعت بیمه در گرو تلفیق داده‌کاوی، هوش مصنوعی و تحلیل پیشرفته داده‌هاست

منابع

احمدی، محمدمهدی (۱۴۰۲). ارائه الگوریتمی مبتنی بر گراف برای کشف تقلب در بیمه درمان، پایان نامه کارشناسی ارشد دانشگاه تهران، پردیس دانشکده‌های فنی – دانشکده علوم مهندسی.

اسدی نژاد، نرگس، عباسی، علی، پورحسین مسعود، یزدانی جمشید (۱۳۹۸). الگوی تصادفات ساختگی در مرکز پزشکی قانونی استان مازندران در سال‌های ۹۵-۱۳۹۰. مجله علمی پزشکی قانونی.. ۲۵(۴ (مسلسل ۹۱)) :۲۰۱-۲۰۷.

اکبرالسادات، سیدمجید و اسماعیل پور قوچانی، بابک (۱۴۰۱). روری بر روشهای کشف تقلب بیمه و بانک با استفاده از هوش مصنوعی، چهارمین همایش ملی مدیریت دانش و کسب و کارهای الکترونیکی با رویکرد اقتصاد مقاومتی، مشهد.

بهرامی، بهاره و محقق، عارفه (۱۴۰۱). زنجیره بلوکی، پاسخگویی و کشف تقلب در شرکت های بیمه، چهارمین همایش ملی تحقیقات میان رشته ای در مدیریت و علوم انسانی.

ذبیحیان، مینا و بیات، نیلوفر (۱۴۰۱). بررسی یک معماری مبتنی بر بلاک چین برای تشخیص تقلب در مدارک درمانی، اولین کنفرانس ملی تحول دیجیتال، بانک و بیمه،تهران.

رحیم خانی، پریسا و منطقی پور، مهناز (۱۴۰۱). مروری بر انواع روش های کشف تقلب مبتنی بر داده کاوی در رشته بیمه های درمان، کنفرانس ملی آخرین دستاوردهای مهندسی داده و دانش و محاسبات نرم،شهرکرد.

رفیق دوست، احمد، مرضیه فریدی، باقری، احمد (۱۴۰۰). تشخیص تقلب بیمه با استفاده از الگوریتم خوشه بندی و شبکه عصبی، اولین کنفرانس ملی بهینه سازی سیستمهای تولیدی و خدماتی،رودسر.

غفاری، مهدی (۱۴۰۲). کاربرد یادگیری ماشینی در تشخیص تقلب؛ مورد بیمه درمان تکمیلی در ایران، پایان نامه کارشناسی ارشد دانشگاه علامه طباطبائی.

فارسیجانی، مهرناز (۱۴۰۱). ارائه مدلی برای شناسایی تقلب در مسیر درمان و تجویز داروهای بیماران کلیوی با استفاده از تکنیک های داده کاوی، پایان نامه کارشناسی ارشد دانشگاه تربیت مدرس، دانشکده مهندسی صنایع و سیستم‌ها.

Amponsah, Anokye Acheampong, Adebayo, Felix Adekoya, Benjamin Asubam, Weyori. (2022). A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology, Decision Analytics Journal 4, 100122. 1-12.

George, K., & Patatoukas, P. N. (2021). The Blockchain Evolution and Revolution of Accounting. In Information for Efficient Decision Making: Big Data, Blockchain and Relevance (pp. 157-172).

Gera, J. R. Palakayala, V. K. K. Rejeti and T. Anusha, (2020). "Blockchain Technology for Fraudulent Practices in Insurance Claim Process,", 5th International Conference on Communication and Electronics Systems (ICCES). 1068-1075

Hamid, Z., Khalique, F., Mahmood, S., Daud, A., Bukhari, A., & Alshemaimri, B. (2024). Healthcare insurance fraud detection using data mining. *BMC Medical Informatics and Decision Making*, *24*. https://doi.org/10.1186/s12911-024-02512-4.

Hong, B., Lu, P., Xu, H., Lu, J., Lin, K., & Yang, F. (2024). Health insurance fraud detection based on multi-channel heterogeneous graph structure learning. *Heliyon*, *10*(9), e30045. https://doi.org/10.1016/j.heliyon.2024.e30045.

Jaiswal, A. Sharma, and S. K. Yadav, K. (2022). Deep feature extraction for document forgery detection with convolutional autoencoders," Comput. Electr. Eng., vol. 99, p. 107770, doi: https://doi.org/10.1016/j.compeleceng.2022.107770.

Kar, A., Navin, L, (2020). Diffusion of blockchain in insurance industry: An analysis through the review of academic and trade literature, Telematics and Informatics, 58(1). 11-16

Lin, Zhao, Huang, L. Z. C., A. et al. (2021). Deep Learning-Based Forgery Attack on Document Images," IEEE Transactions on Image Processing, Volume: 30.

Lu, J., Lin, K., Chen, R. *et al.* (2023). Health insurance fraud detection by using an attributed heterogeneous information network with a hierarchical attention mechanism. *BMC Med Inform Decis Mak* 23, 62 https://doi.org/10.1186/s12911-023-02152-0.

Nabrawi, E., & Alanazi, A. (2023). Fraud Detection in Healthcare Insurance Claims Using Machine Learning. *Risks*, *11*(9), 160. https://doi.org/10.3390/risks11090160.

Naoufel, N. S. W. Syed Sadaf Ali, Iyyakutti Iyappan Ganapathi, Ngoc-Son Vu, Syed Danish A. (2022). Image Forgery Detection Using Deep Learning by Recompressing Images," Electronics, 11(3). 40.

Poddar, J. V. Parikh, and S. K. Bharti, K. (2020). Offline Signature Recognition and Forgery Detection using Deep Learning," Procedia Comput. Sci., vol. 170, pp. 610–617,

Secinaro, S., Dal Mas, F., Brescia, V., & Calandra, D. (2021). Blockchain in the accounting, auditing and accountability fields: a bibliometric and coding analysis. Accounting, Auditing & Accountability Journal

Shaeiri, Z. and Kazemitabar, S. J. (2020). Fast Unsupervised Automobile Insurance Fraud Detection Based on Spectral Ranking of Anomalies. International Journal of Engineering, 33(7). 1240-1248..

Sharma, A. (2020). Analysing the applicability of blockchain accounting and its impact on financial reporting. Sumedha Journal of Management, 9(2). 1-13.

Ramandi, S, Niakan L, Rajaee Harandi S, Asheghi H. (2020). Fraud Detection in Supplementary Health Insurance and Ways to Compete. Iran J Health Insur; 3 (3) :178-187.

Cheng, F., Yan, C., Liu, W., & Lin, X. (2024). Research on medical insurance anti-gang fraud model based on the knowledge graph. *Engineering Applications of Artificial Intelligence*, *134*, 108627. https://doi.org/10.1016/j.engappai.2024.108627.

Li, Jie ; Liu, Jiaying ; Liu, Xin ; Yang, Fang ; Xu, Yong (2022). A medical insurance fraud detection model with knowledge graph and machine learning, Proceedings of the SPIE, Volume 12260, id. 1226023 10 pp.

Hancock, J.T., Bauder, R.A., Wang, H. *et al.* (2023). Explainable machine learning models for Medicare fraud detection. *J Big Data* **10**, 154 https://doi.org/10.1186/s40537-023-00821-5.